PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A1

(11) International Publication Number: **WO 00/34518**

(43) International Publication Date: 15 June 2000 (15.06.00)

(21) International Application Number: PCT/NL99/00743

(22) International Filing Date: 3 December 1999 (03.12.99)

(30) Priority Data:
98204114.7 4 December 1998 (04.12.98) EP

(71) Applicant (*for all designated States except US*): KEYGENE N.V. [NL/NL]; P.O. Box 216, NL–6700 AE Wageningen (NL).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): VOS, Petrus, Antonius, Josephina [NL/NL]; Dorpsstraat 22, NL–3927 BD Renswoude (NL). VAN EIJK, Michael, Josephus, Theresia [NL/NL]; Tolsteegplantsoen 43, NL–3523 AN Utrecht (NL). HOGERS, René, Cornelis, Josephus [NL/NL]; Aalbeek 38, NL–6715 GR Ede (NL). HEIJNEN, Leo [NL/NL]; Kerkstraat 30 A, NL–4132 BE Vianen (NL).

(74) Agent: DE BRUIJN, Leendert C.; Nederlandsch Octrooibureau, Scheveningseweg 82, P.O. Box 29720, NL–2502 LS The Hague (NL).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ARRAY AND METHOD FOR ANALYSING NUCLEIC ACID SEQUENCES

(57) Abstract

The invention relates to an array for analysing a nucleic acid sequence or mixture of nucleic acid sequences, comprising: a) a carrier; and b) at least two different nucleic acid sequences bound to said carrier, in which each of the nucleic acid sequences bound to the carrier comprises at least a nucleic acid sequence that corresponds to the sequence of a restriction fragment obtainable by restricting a genomic DNA and/or at least one cDNA with at least one frequent cutter restriction enzyme and at least one rare cutter restriction enzyme. The array preferably comprises at least 10, preferably at least 100, more preferably at least 1000 different nucleic acid sequences bound to the carrier. More preferably, the array comprises a plurality of AFLP®–markers taken from a single individual or from a group of related individuals. The invention also relates to a method for providing nucleic acid sequences, and in particular AFLP®–markers, for use in such an array, as well as to a method for analysing a nucleic acid sequence or a mixture of nucleic acids sequences using such an array. Finally, the invention relates to a kit of parts comprising an array of the invention, as well as to data obtained using the array of the invention.

Array and method for analysing nucleic acid sequences.

The present invention relates to arrays for analysing nucleic acid sequences and to methods for analysing nucleic acid sequences using such an array.

In particular, the invention relates to arrays and methods for determining whether a specific nucleic acid sequence is present or absent in a nucleic acid sequence or mixture of nucleic acid sequences.

More in particular, the invention relates to an array and a method for determining the presence or absence, in genomic DNA or a sample of restriction fragments derived from genomic DNA, of sequences corresponding to unique restriction fragments that can serve as genetic markers, such as AFLP-markers.

The invention further relates to a method for preparing such an array, in particular in the form of a high density array for the detection of biological molecules, herein referred to as a "biochip".

A number of methods for analyzing nucleic acid sequences are known. In general, these methods comprise immobilization of the sequences to be analysed, for instance by blotting; hybridization of the sequences with a labeled DNA- or RNA-probe; stringency washes to remove non-hybridized material; followed by detection of those sequences that have hybridized with the probe.

Such techniques are often carried out after prior amplification -such as by PCR- of the starting nucleic acid sequences, usually a mixture of restriction fragments from a genomic DNA. The resulting mixture of amplified fragments is then separated, for instance on the basis of differences in length or molecular weight, such as by gel-electroforesis, and then visualised, i.e. by blotting followed by hybridization. The resulting pattern of bands is referred to as a DNA fingerprint.

Usually in DNA fingerprinting, fingerprints of closely related species, subspecies, varieties, cultivars, races or individuals are compared. Such related fingerprints can be identical or very similar, i.e. contain a large number of corresponding -and therefore less informative- bands.

Differences between two related fingerprints are referred to as "DNA polymorphisms". These are DNA fragments (i.e. bands) which are unique in or for a

primers hybridized; and

(e) identifying or recovering the amplified or elongated DNA fragment thus obtained.

The thus amplified DNA-fragments can then be analysed and/or visualised, for instance by means of gel-electrophoresis, to provide a genetic fingerprint showing bands corresponding to those restriction fragments that have been linked to the adapter, recognized by the primer, and therefore amplified during the amplification step.

The AFLP-fingerprint thus obtained provides information on the specific restriction site pattern of the starting DNA. By comparing AFLP-fingerprints from related individuals, bands which are unique for each fingerprint can be identified. These polymorfisms are referred to as "AFLP-markers", and can again be used to identify a specific individual, cultivar, race, variety, subspecies or species, and/or to establish the presence or absence of a specific inherited trait, gene or disease state.

AFLP thus requires no prior knowledge of the DNA sequence to be analysed, nor prior identification of suitable probes and/or the construction of a gene library from the starting DNA.

For a further description of AFLP, its advantages, its embodiments, as well as the techniques, enzymes, adapters, primers and further compounds and tools used therein, reference is made to EP-A-0 534 858 and co-pending European applications 98.202.5496 and 98.202.4515, all by applicant and incorporated herein by reference. Also, in the description hereinbelow, the definitions given in paragraph 5.1 of EP-0 534 858 will be used, unless indicated otherwise.

Although AFLP is generally less time-consuming than hybridisation-based techniques, it still suffers from the disadvantage that the amplified fragments have to be separated (i.e. by gel-electrophoresis) and visualized (i.e. by generation of a fingerprint). These are very elaborate and time consuming procedures, which require special apparatus, such as electrophoresis and auto-radiography equipment. Thereafter, the fingerprints have to be analysed -nowadays generally performed by "reading" the fingerprint into a computer- to identify the polymorphic bands. Generally, this also requires to use of a known reference sample run at the same time in a parallel lane of the gel.

According to the invention, this is carried out by analysing the genomic DNA of two or more related individuals using AFLP, identifying polymorphisms ("AFLP-markers") within the genome, amplifying and isolating the nucleic acid sequences corresponding to these AFLP-markers, and binding the amplified sequences to specific areas of a carrier, thus providing an array comprising essentially only nucleic acid sequences that correspond to AFLP-markers.

This array can then be used to analyse a sample of nucleic acid(s) -such as a genomic DNA or restriction fragments thereof- derived from the same or a genetically related individual, by contacting the sample with the array under hybridizing conditions. The nucleic acid sequence(s) to be analysed will then (only) hybridize with those parts of the array that carry an essentially homologous sequence, i.e. the same AFLP-marker, or at least a sequence with a high degree of homology with the marker. Thus, by analysing to which parts of the array (i.e. to which AFLP-markers) the nucleic acid sequence(s) to be analysed has or have hybridized, the presence of absence of said marker in the sample can be established.

In other words, the invention makes it possible to test a sample of nucleic acid(s) directly for the presence of a large number of polymorphic fragments or bands - i.e. as many as are bound to the carrier- without the need of generating and analysing a DNA-fingerprint.

The invention also makes it possible to test simultaneously for a large number of "unrelated" markers (i.e. markers which can normally not be detected in a single AFLP-reaction or fingerprint) by incorporating these different markers into a single array.

Other objects and advantages of the invention will become clear from the description hereinbelow.

H. Himmelbauer et al., Mammalian Genome 9, 611-616 (1998) describe a method for the identification and mapping of polymorphic markers, using "*a modification of the AFLP technique*" called the "*IRS-PCR system*". According to this method, genomic (mouse) DNA is restricted using a single restriction enzyme (*SacI* or *Bam*HI), amplified in a PCR using adapters and primers, after which the amplicons thus obtained are hybridized with a gridded genomic library (BAC-clones) to identify

usually by comparison to known results or a reference using sophisticated computer algorithms.

In a first aspect, the present invention relates to an array for analysing a nucleic acid sequence or a mixture of nucleic acid sequences, comprising:

a) a carrier; and

b) at least two different nucleic acid sequences bound to said carrier, in which each of the nucleic acid sequences bound to the carrier comprises at least a nucleic acid sequence that corresponds to the sequence of a restriction fragment obtainable by restricting a genomic DNA with at least one frequent cutter restriction enzyme and at least one rare cutter restriction enzyme.

More particular, the invention relates to such an array in which at least 50%, preferably at least 70%, more preferably at least 90% of the nucleic acid sequences bound to the carrier comprise the sequence of a restriction fragment that corresponds to an AFLP-marker.

In a further aspect the invention relates to a method for providing an array of nucleic acid sequences bound to a carrier, comprising the steps of:

a) identifying an AFLP-marker;

b) providing a nucleic acid sequence that comprises a restriction fragment sequence corresponding to said AFLP-marker;

c) attaching the nucleic acid sequence to the carrier; and

d) repeating steps a) to c) for different AFLP markers to build up an array.

More particularly, the invention relates to such a method comprising the steps of:

a) identifying a polymorphic band in an AFLP-fingerprint;

b) isolating a nucleic acid sequence from said polymorphic band;

c) optionally further amplifying, purifying and/or modifying the nucleic acid sequence; and

d) attaching the nucleic acid sequence to the carrier.

e) repeating steps a) to d) for different polymorphic bands to build up an array.

In a yet another aspcect, the starting DNA used to generate the restriction fragments that are bound to the carrier are not derived from genomic DNA, but from at

In the description, the nucleic acid sequences bound to the carrier will be indicated as "*Array-bound Nucleic Acid Sequence(s)*" or "ÄNAS", and the restriction fragments present therein will be indicated as "*Restriction Fragment Sequence(s)*" or "RFS". Usually, each Array-bound Nucleic Acid Sequences will comprise (only) one Restriction Fragment Sequence, and optionally further nucleic acid sequences or structural elements as described below, bound to the Restriction Fragment Sequence. When Array-bound Nucleic Acid Sequences are referred to hereinbelow as "different", it means that these Array-bound Nucleic Acid Sequences contain different Restriction Fragment Sequences.

The array preferably comprises at least 10, more specifically at least 100, more preferably at least 1000 different Array-bound Nucleic Acid Sequences. For a "high-density array" or "micro-array", the total number of Array-bound Nucleic Acid Sequences will be in the region of 100 - 100.000.

These Array-bound Nucleic Acid Sequences will generally be bound to the carrier in such a way that each Array-bound Nucleic Acid Sequence is attached to, and corresponds with, a specific, distinct part of the carrier, so as to form an independently detectable area on the carrier, such as a spot or band. This makes it possible to "read" the array by scanning (i.e. visually or otherwise) the areas to which the Array-bound Nucleic Acid Sequence (i.e. the marker) of interest is attached.

Preferably, the Array-bound Nucleic Acid Sequences are bound to the carrier in accordance with a predetermined, regularly distributed pattern, in which for instance related Array-bound Nucleic Acid Sequence (i.e. related markers) can be grouped together, i.e. in one or more lines, columns, rows, squares, rectangles, etc, preferably in an "adressable" form. This further facilitates analysis of the array.

The density of the different Array-bound Nucleic Acid Sequences will generally be in the region of 1-100,000 different markers/cm^2, usually 5-50,000 markers/cm^2, generally between 10-10,000 markers/cm^2.

In general, each of the Array-bound Nucleic Acid Sequences on the array will correspond to a specific polymorphic band or marker, i.e. as derived from an AFLP-fingerprint of genomic DNA of a specific individual. Usually, the array will comprise sets of one or more of such markers taken from a single fingerprint, or at least taken

In one embodiment, the AFLP-markers present on the array have been taken from or will be representative of different subspecies, varieties, cultivars, lines or races of the same species.

An array of the invention can also contain markers representative of a certain genetic state of an individual, such as the presence or absence of a disease state, i.e. of oncogenes and of genetically determined diseases.

As already mentioned above, besides arrays based on restriction fragments derived from genomic DNA - e.g. based on polymorphic fragments/genetic markers - the invention also provides arrays based on (restriction fragments derived from) cDNA.

According to this aspect of the invention, the RFS present in the ANAS will be a restriction fragment obtained by restricting at least one cDNA with at least one restriction enzyme, and preferably with at least one frequent cutter restriction enzyme and at least one rare cutter restriction enzyme as described herein.

Usually, prior to attachment to the array, the cDNA-derived restriction fragments thus obtained are amplified, preferably using AFLP. Such AFLP-amplification of cDNA is generally referred to as "cDNA-AFLP" and can be carried out essentially as described above for the AFLP-amplification of genomic DNA and/or by using any cDNA-AFLP protocol known per se, to provide a cDNA-derived AFLP-fingerprint.

One or more of the bands from this cDNA-AFLP fingerprint may then be isolated from the gel and bound to the array, e.g. after re-amplification and/or incorporation into an ANAS, essentially as described for the genomic DNA.

This may be carried out for different bands obtained from the same cDNA, and/or for bands from one or more different cDNAs. Also, the one or cDNAs used to provide the RFS may be obtained from (mRNAs derived from) one individual (e.g. from different cells, parts, tissues or organs) and/or from two or more individuals, e.g. individuals belonging to same race, variety, species, genus, family etc.., with the same or different phenotypical characteristics. Also, the cDNAs may be obtained from (mRNA derived from) healthy individuals and/or from diseased individuals; and/or from individuals at different stages of development.

Furthermore, although the genomic DNA based arrays and the cDNA based

also makes it possible to identify and prepare, at the same time, markers from related individuals (i.e. from one genotyping collection) i.e. by running parallel AFLP-reactions and visualizing these reactions in adjacent lanes of the same gel. In this way, a micro-array containing a large number of markers and/or containing all relevant markers from a genotyping collection can be build up very efficiently.

As in AFLP, two different restriction enzymes are used to digest the starting (genomic) DNA, i.e. the "frequent cutter", which serves the purpose of reducing the size of the restriction fragments to a range of sizes which are amplified efficiently, and the "rare cutter" which serves the purpose of targeting rare sequences. For both, reference is made to EP-A-0 534 858 and EP-A-0 721 987 by applicant, incorporated herein by reference.

Examples of suitable frequent cutter enzymes are *Mse*I and *Taq*I. Examples of commercially available rare cutters are *Pst*I, *Hpa*II, *Msp*I, *Cla*I, *Hha*I, *Eco*RII, *Bst*BI, *Hin*P1, *Mae*II, *Bbv*I, *Pvu*II, *Xma*I, *Sma*I, *Nci*I, *Ava*I, *Hae*II, *Sal*I, *Xho*I and *Pvu*II, of which *Pst*I, *Hpa*II, *Msp*I, *Cla*I, *Eco*RII, *Bst*BI, *Hin*P1 and *Mae*II are preferred.

The AFLP-reaction will usually be carried out following known protocols, for which reference is made to EP-A-0 534 858, incorporated herein by reference.

The Restriction Fragment Sequence (with the AFLP-adapters) will generally have a size that can be detected as an individual band in an AFLP-fingerprint, i.e. in the range of 50 – 1200 base pairs. It will be clear that, as the Restriction Fragment Sequence are separated by gel-electrophoresis, they will be of different sizes.

Also, it may be possible to use as the Restriction Fragment Sequence only a part of a restriction fragment obtained as/from a band in the AFLP fingerprint. Such a partial sequence may for instance be obtained by (further) restricting the restriction fragment(s) isolated from the AFLP gel with one or more restriction enzymes, i.e. usually with other restriction enzymes than the one or two orginally used to generate the restriction fragments from the starting genomic or cDNA, including but not limited to synthesized oligonucleotides based and/or derived thereof. For this purpose, any desired and/or pre-determined restriction enzyme or enzyme combination may be used; suitable restriction enzymes include, but are not limited to, the frequent cutters and rare cutters mentioned above, IIS-type restriction enzymes.

Fragment Sequence itself, depending upon the technique used for binding the Array-bound Nucleic Acid Sequence to the array, as further described below.

The carrier for the array may be any solid material to which nucleic acid sequences can be attached, including porous, fibrous, woven and non-woven materials, as well as composite materials. Also, semi-solid materials such as gels or matrices (for instance as used in chromatography) may be used, although this is not preferred.

Suitable carriers include, but are not limited to, those made of plastics, resins, polysaccharides, silica or silica-based materials, functionalized glass, modified silicon, carbon, metals, inorganic glasses, membranes, nylon, natural fibers such as silk, wool and cotton, and polymer materials such as polystyrene, polyethylene glycol tetraphthalate, polyvinyl acetate, polyvinyl chloride, polyvinyl pyrrolidone, polyacrylonitrile, polymethyl methacrylate, polytetrafluoroethylene, butyl rubber, styrenebutadiene rubber, natural rubber, polyethylene, polypropylene, (poly)tetrafluoroethylene, (poly)vinylidenefluoride, polycarbonate and polymethylpentene. Further suitable support materials are mentioned for instance mentioned in US-A-5,427,779, WO 97/22720, WO 97/43450, WO 97/31256, WO 97/27317 and EP 0 799 897.

Preferred carrier materials are glass and silicon.

Preferably, the carrier will have an essentially flat, rectangular shape, with the Array-bound Nucleic Acid Sequences bound to one surface thereof. However, any other suitable two- or three-dimensional form may also be used, such as a disc, a sphere or beads, or materials or structures that allow a liquid medium containing the sample to be analysed to pass or flow through the carrier, such as columns, tubes or capillairies, as well as (macro)porous-, web- or membrane-type structures, including the flow-through genosensor devices referred to in WO 97/22720.

The size of the array, as well as of the individual areas corresponding to each of the different Array-bound Nucleic Acid Sequences, may vary, depending upon the total amount of Array-bound Nucleic Acid Sequence, as well as the intended method for analysing the array.

For an array that is to be inspected visually, the total array and the separate areas thereon will be of such a size that they can be seen and distinguished with the naked eye or through a microscope, i.e. in the range of 1 to 500 cm^2 for the total array,

or epoxysilane-amine chemistry, for instance as described in WO 97/22720, also incorporated herein by reference.

An example of a strong, but non-covalent binding technique involves the attachment of a biotinylated Array-bound Nucleic Acid Sequence onto a carrier coated with streptavidin.

In order to create small, distinct, adressable areas of each of the Array-bound Nucleic Acid Sequence on the array, masking techniques or known microdispensing techniques may be used, for instance as described in WO 97/46313 and WO 97/22720.

After attachment of the Array-bound Nucleic Acid Sequences to the carrier, the array will generally be ready for use.

In a further aspect, the invention relates to a method for analysing a nucleic acid sample using the array of the invention. In general, this method comprises contacting the sample to be analysed with the array under hybridizing conditions, so that the one or more of the nucleic acid sequence(s) present in the sample may bind to the one or more of the Array-bound Nucleic Acid Sequences on the array, more specifically with the Restriction Fragment Sequences present in the Array-bound Nucleic Acid Sequence. This method is described in more detail in the Experimental Part below.

Usually, a nucleic acid sequence or mixture will be analysed that is suspected to comprise at least one sequence or fragment that corresponds to a Restriction Fragment Sequence (i.e. an AFLP-marker) present on the array used. In this context, "corresponds" means a sequence homology of at least 70%, more preferably at least 85%, specifically 95%-100%.

In general, the method of the invention is based on the hybridisation of sequences in the sample to be analysed with the Restriction Fragment Sequence. In other words, in the invention, the target sample is probed directly with the pre-selected sequences/markers of interest, so that a positive hybridization event or signal is directly indicative of the presence of said marker in the target sample. Also, as these markers are unique sequences with low abundance in the target genome, generally a high selectivity can be obtained,

Also, in a highly preferred embodiment of the invention, in analysing a target

40°C to 1X SSPE-T at 37°C down to as low as 0.25X SSPE-T at 37-50°C.

The hybridisation conditions are preferably chosen such that only those nucleic acid sequences in the target sample that have more than 70%, preferably more than 80%, more preferably more than 90% homology, and in particular 95-100% homology with the Restriction Fragment Sequences, will hybridize with the Array-bound Nucleic Acid Sequence. These will generally be "moderate" or preferably "stringent" hybridisation conditions. Such stringent conditions can be as described in EP 0 799 897.

After hybridization, the array is washed to remove unwanted compounds, in particular any nucleic acid sequences not hybridized with the Array-bound Nucleic Acid Sequences on the array. Thereafter, the array is analysed to determine to which areas on the array (i.e. to which Array-bound Nucleic Acid Sequences/Restriction Fragment Sequences) the nucleic acid sequence(s) from the sample has/have hybridized. These area's will generally be detected as a positive signal indicating the presence of the marker in the sample.

The analysis of the array may be carried out in any manner known per se, including optical techniques, spectroscopy, chemical techniques, biochemical techniques, fotochemical techniques, electrical techniques, light scattering techniques, colorimetric techniques, radiography techniques, etc., as long as they can indicate the presence of a hybridization event. Suitable techniques are for instance described in WO 97/27317, WO 97/22720, WO 97/43450, EP 0 799 897, WO 97/31256, WO 97/27317 and WO 98/08083.

Usually, a technique using detectable labels will be used. Such a label will generally be attached to the nucleic acid sequence(s) to be analysed, so that -after hybridization with the array- those areas of the array which show the presence of the label correspond to a positive hybridization event.

Suitable labels are for instance described in WO 97/27317, WO 97/22720, WO 97/43450, EP 0 799 897, WO 97/31256, WO 97/27317 and WO 98/08083 and include fluorescent labels, phosphorescent labels, chemoluminescent labels, bioluminescent labels, chemical labels, biochemical labels such as enzymes, biological labels such as biotin/streptavidin, radioisotopes, spin or resonance labels, metal colloids

comprises steps (a) - (e) of the general AFLP-method described above, in which step (e) is carried out by contacting the (mixture of) amplified or elongated DNA fragment(s) obtained in step (d) with an array as described herein.

Compared to conventional fingerprinting/autoradiography, the use of an array generally will be faster than using fingerprinting/autoradiography, and several markers that would require generating several separate fingerprints could be combined into a single array. This makes the arrays of the invention especially suited for routine and/or high throughput screening, for instance in plant breeding.

Also, the array of the invention can conveniently be provided as a kit of parts comprising the array and other components for use with the array, such as restriction enzymes, polymerase(s), adapters, primers, buffers, nucleotides, labels or other detection agents, containers/packaging and manuals. The array of the invention may even be in the form of a hand-held device such as a dipstick.

The array of the invention may be re-usable, usually through regeneration to remove the hybridized sequences. A kit of the invention may therefore also contain agents that can be used for such regeneration.

The array of the invention can be used to analyse any kind of nucleic acid sequence or mixture of nucleic acid sequences, including, but not limited to, plant-derived sequences, animal-derived sequences, human-derived sequences, microbial sequences, yeast sequences, sequences from fungi and algi, and viral sequences, depending upon the origin of the restriction fragment sequences bound to the array, including but not limited to whether the restriction fragments bound to the array are derived from genomic DNA or cDNA (or both).

Also, the array may be used to analyse DNA sequences, including genomic DNA, cDNA, structural genes, regulatory sequences and/or parts thereof; as well as RNA, including mRNA, optionally by analogous modification of the method given above.

The nucleic acid sample analysed with the array may be a sample as isolated directly from a living or dead organism or from tissue or cells. Preferably however, prior to hybridisation with the array, the nucleic acid sample is restricted with one or more restriction enzymes, preferably the same two restriction enzymes used in

invention can be used for any purpose for which the use of cDNA-AFLP is envisaged, including but not limited to applications such as expression profiling, functional genomics, and gene mapping. For any of these applications, it is envisaged that - as with cDNA AFLP - a cDNA-based array may be used to determine both qualitatively as well as quantitatively - e.g. based on the strength of the hybridisation signal obtained with the array - the presence of one or more specific nucleotide sequences in a starting sample. These may include both DNA-sequences as well as RNA-sequences, including expresssion-dependant RNA sequences such as mRNAs.

Possible fields of use of both the genomic DNA-based as well as the cDNA-based arrays are for instance plant and animal breeding, variety or cultivar identification, diagnostic medicine, disease diagnosis in plants and animals, identification of genetically inherited diseases in humans, family relationship analysis, forensic science, organ-transplant, microbial and viral typing such as multiplex testing for strains of infectious diseases; as well as the study of genetic inheritance, gene expression, mutations, oncogenes and/or drug resistance; or for mRNA detection.

Arrays of the invention may further be developed for and used in any other application for which known nucleotide arrays are used or envisaged. These include the applications mentioned in for example WO 97/27317, WO 97/22720, WO 97/43450, EP 0 799 897, EP 0 785 280, WO 97/31256, WO 97/27317 and WO 98/08083.

As already mentioned above, in these applications, it is envisaged that arrays of the invention can be developed that carry most or even all markers of interest for a specific genotyping collection, such as for a specific species. Other arrays of the invention may contain most or all markers necessary to classify an individual within a genotyping collection, i.e. as belonging to a certain species, subspecies, variety, cultivar, race, strain or line, or to study the inheritance of a genetic trait or property. Also, an array of the invention may contain all markers indicative of the presence, the absence or the state of a genetically determined or genetically influenced disease or disorder, including cancer, oncogenes and oncogenic mutations. Such an array may then be used for diagnostic purposes.

Similarly, it may be possible to provide cDNA-based arrays for any of these purposes.

tomato cDNA-AFLP fragments.

- Figure 8: Detection of rice AFLP markers amplified in a +2/+3 AFLP reaction on an array containing 5 rice +2/+3 AFLP markers and 5 sets of oligo's corresponding to the forward and reverse strands of these 5 rice +2/+3 AFLP markers.
- Figure 9: Schematic representation of the related AFLP-Primer Combinations ("ÄPCs") used in Example I. Fig. 9A: 4 related APCs of the Enzyme Combination ("EC") *Eco*RI-*Mse*I; Fig. 9B: APC that can be used to simultaneously amplify the 4 APCs of Fig. 9A; Fig. 9C: 16 related APCs of the EC *Pst*I-*Taq*I; Fig. 9D: APC that can be used to simultaneously amplify the 16 APCs of Fig. 9C.
- Figure 10: Schematic representation of the method used in generating the Array-bound Nucleic Acids Sequences (including the AFLP-amplification);
- Figure 11: Schematic representation of the method for identifying, in/from a plurality of AFLP-fingerprints, polymorphic bands suitable for use in an array of the invention, and for "building up" an array from such polymorphic bands;
- Figure 12: Schematic representation of a method for probing a genomic DNA with an array of the invention, in which the genomic DNA is restricted and amplified using AFLP-methodology prior to contacting with the array.

Experimental Part.

Example I: Generation of AFLP micro-arrays.

The method for generating the AFLP micro-arrays comprises steps (1) - (9). Steps (2) - (5) generally follow conventional AFLP-techniques and protocols, as described in EP-0 534 858. A number of steps of said method, as well as the primers/primer combinations used therein, are schematically shown in Figures 9-11.

1. Selection of the genotyping collection.

A limited number of individuals is selected representing the genetic diversity within a specific group in the best possible way. The selected set of individuals is called the "genotyping collection".

Preferably, a large number of AFLP reactions is performed on the genotyping collection using a set of "related AFLP-Primer Combinations", hereinbelow referred to as "related APCs". The APCs mentioned in this Example are also schematically shown in Figures 9A-9D.

Related APCs are combinations of selective AFLP-primers that can be used with the same Restriction Enzyme Combination and that can be amplified simultaneously using a corresponding APC with less selective nucleotides, yielding all AFLP fragments from the related APCs at once. Each of the primers that forms part of an APC is essentially the same as a conventional AFLP-primer in that it contains:

1) a sequence corresponding to (i.e. that can hybridize with) the adapter-sequence of the template, connected at its 3' end to:

2) a (usually small) sequence that corresponds to the part of the template sequence that resulted from the cutting of a restriction site in the original genomic DNA with the restriction enzyme used and the ligation of the restricted fragment to the adapter; and

3) at the 3' end of the primer, a number of so-called selective bases, for which further reference is made to EP-0 534 858.

Each primer of an APC can be represented schematically as:

5'- AAAAAAAAAAAAA - RRR - NNN -3'

in which N is a nucleotide corresponding to the adapter sequence, R is a nucleotide corresponding to the restriction sequence, N is a selective nucleotide (the number of nucleotides A, R, and N may vary and may be different than shown); or alternatively as

[adapter] - [restr.enzyme] - NNN

in which [adapter] is the adapter sequence, [restr.enzyme] is the restriction sequence, and N is a selective nucleotide.

Each APC will consist of two primers, i.e. one primer for the rare cutter and one primer for the frequent cutter. A set of APCs will comprise a number of such two-primer APCs.

To provide a set of "related APCs", the last selective base at the 3' end of the primer for the frequent cutter, of the primer for the rare cutter, or of the primers for both, may be varied to two or more, and preferably all four of the bases A, T, G and C.

[Adapter] - [*Pst*I] - CT + [Adapter] - [*Taq*I] - AGG

[Adapter] - [*Pst*I] - CT + [Adapter] - [*Taq*I] - AGT

These APCs can be used with the Restriction Enzyme Combination *Pst*I-*Taq*I and have the selective nucleotide C at the *Pst*I-primer and the selective nucleotides AG at the *Taq*I-primer in common. The AFLP-fragments from these 4 related APCs can be amplified at once with the APC

[Adapter] - [*Pst*I] - C + [Adapter] - [*Taq*I] – AG (Figure 9D)

Preferably, in an APC, primers with 1, 2, 3 or 4 selective nucleotides are used. More preferably, each APC comprises a combination of two +3-primers, or one +3-primer and one +2-primer, or two +2 primers.

4. Fingerprinting.

After a suitable set of related APCs has been selected, the restricted genomic DNA of an individual from the genotyping collection is amplified using one APC from the set.

This is carried out for all individuals of the genotyping collection, in separate amplifications, carried out essentially as described in EP-0 534 858, which are usually run simultaniously.

The resulting AFLP reactions, one for each individual of the genotyping collection, are then analyzed in parallel on sequencing gels. After electrophoresis, these gels are dried on Whatman 3 MM paper and the AFLP fingerprints are visualized, e.g. by autoradiography or phospho-imaging.

In this way the AFLP fingerprints of the individuals of the genotyping collection are displayed side by side on the fingerprint. This is schematically shown in Figure 11, in which the AFLP-reactions of a genotyping collection of 4 individuals (referred to "ind.1" to "ind.4" – and corresponding to the lanes from left to right in the gels) have been visualised in the four parallel lanes of each of the gels "pk 1" to "pk 4" (in which each gel was generated with a different APC from a set of four related APCs).

5. Identification of polymorphic bands.

EXAMPLE II: Carrier attachment and formation of an array.

The individual AFLP markers of the AFLP fragment library are attached to a carrier; many different AFLP fragments are attached to the same carrier. This is preferably carried out according to a predetermined pattern, in which for instance the markers generated from the genotyping collection with a specific APC are grouped together, i.e. as a column as shown in Figure 11.

Also, the markers generated with each of the APCs from the set of related APCs may be grouped together, to form a set of lines, rows or columns, or a rectangle, as is shown in Figure 11.

In this way an array of AFLP markers is created on the carrier. In case of a high-density array such arrays are called AFLP micro-arrays. Usually, each APC will provide about 10-50 markers, depending upon the genotyping collection and the number of individuals used.

The array thus obtained can then be used to probe the genomic DNA of a further individual for the presence of the AFLP markers attached to the array, as further described in Example III. Usually, this further individual will belong to or be related to the genotyping collection used in generating the array, or at least will be suspected of containing in its genome one or more of the markers present on the array.

EXAMPLE III: Genotyping using AFLP micro-arrays.

This procedure uses the AFLP micro-arrays obtained as described in Example I. Such micro-arrays contain a multitude of AFLP markers derived from a specific genus. (In general AFLP markers will be genus-specific and AFLP markers generated from a different genus will usually not be usable for genotyping of individuals from other genera).

Genotyping of a specific individual can be performed by investigating the presence or absence of each AFLP marker of the AFLP micro-array in the individual tested. This can for instance be achieved by hybridization of a collection of AFLP-fragments from the individual to the AFLP markers attached to the micro-array.

This collection of AFLP fragments is preferably generated from the individual of interest by AFLP amplification of AFLP template DNA of the individual. The

less selective nucleotides than the primers of the set of related APCs.

Usually, the primers of said APC will contain no selective bases on the positions varied in the primers of the set of "related APCs" used in generating the array, as exemplified in step 3 of Example I above. The remainder of the selective bases in primers of said APC will be the same as in the primers of the set of related APCs, also as exemplified in step 3 of Example I.

In principle, using said APC, all fragments that have been amplified separately with the set of related APCs can be amplified together. Therefore, by using said APC on the template DNA of step 1, a mixture of amplified fragments can be generated that will contain any marker generated with the set of related APCs, if such a marker is/was also present in the genomic DNA to be tested.

For the remainder, the amplification is carried out in a manner known per se, for instance essentially as described in EP-0 534 858, and preferably in a manner analogous to step 3 of Example I, i.e. following the same or a similar protocol.

During or after amplification, the AFLP fragments are labeled by using end-labeled AFLP primers, or by internal labeling. The label may be a fluorescent label, a radio-active label, or other types of labels suitable for detection on micro-arrays.

3. Hybridization with the array.

The labeled AFLP fragments generated with the selected APC are used as a probe in a hybridization to the AFLP fragments on the AFLP micro-array. The collection of labeled AFLP fragments is called the "AFLP target". AFLP markers represented on the AFLP micro-array will hybridize to their labeled counterparts in the AFLP target, provided that these AFLP markers are present in the individual selected. The result is that the AFLP markers on the array that correspond to markers present in the individual tested will hybridize to their labeled counterparts, and give a positive hybridization signal on the array (i.e. show the presence of the label).

AFLP markers on the array that are not present in the individual tested will not find corresponding labelled sequences in the amplified sample, and will therefore not give a positive signal.

2. Reamplification

*Eco*RI+A/*Mse*I+C fragments are reamplified with the following primers 98L19 and 98L20, that reconstitute the restriction sites.

98L19:AGCGGATAACAATTTCACACAGGATAGACTGCGTACGAATTCA
M13 reverse sequence primer X
E01: GACTGCGTACCAATTCA

98L20:CGCCAGGGTTTTCCCAGTCACGACGATGAGTCCTGATTAAC
M13 forward sequence primer X
M02: GATGAGTCCTGAGTAAC

(98L19 = SEQ ID no.1 ; E01 = SEQ ID no.2 ; 98L20 = SEQ ID no. 3; M02 = SEQ ID no.4)

The PCR reaction mixture is as follows: 5 µl eluate; 150 ng 98L19; 150 ng 98L20; 2 µl 5 mM dNTP's; 5 µl 10 x PCR buffer; 0.2 U Taq polymerase, in a total volume of 50 µl. The PCR profile is as follows: 30 sec. 94°C : 30 sec. 56°C : 1 min. 72°C, for 30 cycles.

*Pst*I+A/*Mse*I+C fragments are reamplified with primers 98/L88 and 98/L20

98L88:AGCGGATAACAATTTCACACAGGATAGACTGCGTACCTGCAGA
M13 reverse sequence primer X
P01 GACTGCGTACATGCAGA
X
Ps01 GACTGCGTACCTGCAGA

(98L88 = SEQ ID no.5 ; P01 = SEQ ID no.6 ; Ps01 = SEQ ID no.7) or optionally with 98L89/98L20.

98L89:AGCGGATAACAATTTCACACAGGATAGACTGCGTACCTGC
M13 reverse sequence primer X
P00: GACTGCGTACATGCAG

(98L89 = SEQ ID no.8 ; P00 = SEQ ID no.9)

5. Reamplification and validation of the cloned fragments

The following reamplification primers are used:

- rare cutter side: 98L58: GGAAACAGCTATGACCATGATTAC (pUC 18 primer, SEQ ID no.10)
- *Nde*I side: 98L55 GATTGTACTGAGAGTGCACCTTAAC (pUC 18 primer, with reconstituted *Mse*I site, only for Mse+C, SEQ ID no.11).

For each clone fragment 3 different clones and inoculated into 10 µl TY, followed by incubation (o/n) at 37°C. *E. coli* cells are transferred to 96-wells plate with 50 µl $TE_{0,1}$ per well, and 5 µl is transferred to PCR base. The PCR base is incubated at 95°C during 5 min, after which 45 µl PCR mixture is added, which comprises: 75 ng primer 98L58; 75 ng primer 98L55; 2 µl 5 mM dNTP's; 5 µl 10X PCR buffer; 0.25 µl Taq polymerase; 0.85 µl 10 mg/ml BSA, to a total volume of 45 µl. The PCR profile is as follows: 25 sec. 94°C; 30 sec. 56°C; 1 min. 72°C; for 30 cycles. 5 µl of the mixture is checked on the gel.

For each APC from which a fragment is obtained, 3 pools are made. Pool A contains fragment 1, 4, 7, 10 ...; Pool B contains fragment 2, 5, 8, 11 ...; Pool C contains fragment 3, 6, 9, 12 .... For each clone 5 µl colony PCR material is pooled, and 5 µl of each pool us used for a template preparation (standard AFLP template).

The template is checked by standard AFLP reaction of 1/10 diluted pool template, the fingerprint of which is compared to the fingerprint from which the original fragments were obtained.

EXAMPLE V: Protocols for detecting AFLP fragments using micro arrays.

Preparing the micro arrays

The micro arrays are prepared using DNA probes that are synthesized via "colony PCR" using pUC18 specific primers. For preparing the arrays, DNA solutions at a concentration of about 0.5 µg/µl are used. Diluted colony PCR material is used for routine synthesis of probe DNA.

1. Amplification of the probes

To synthesize as much DNA as possible in a small a volume as possible, an

and serves as a control to monitor deposition and binding to the array.The slides used were EMS Poly-L-lysine slides (Electron Microscopy Scienses, Washington).
Printing of the slides is carried out as follows:

a) Position of the slides:

- Piece of "matted glass" on the slides on the left side (against the clamp).
- Press slides well together.

b) Position of the microtiter plates:

- A1 coordinate of the plates left side front in the plate holders.

c) Settings arrayer:

1x spotting X = 3 Y = 15, duplo X = 7.4 Y = 15

5x spotting X = 3 Y = 19, duplo X = 7.4 Y = 19

spot spacing: 300 μm - 350 μm.

To limit the evaporation of the probe DNA's, the microtiter plates can be kept above a bath of warm water.

4. Processing of the arrays

During the processing of the arrays, DNA is adhered to the glass carrier and denaturated, depending upon the type of slide and the coating. The processing of EMS Poly-L-lysine slides is carried out as described by P. Brown (http://cmgm.stanford.edu/pbrown/protocols/3_post_process.html):

The slide is rehydrated on top of a hot water bath for 1 minute, so that the slide becomes fogged, and snap dried on a heated cooking plate (about 3 sec.). The slide is then rehydrated for 10 sec, UV cross-linked at 65 mJ (Amersham UV crosslinker at 650 x 100 μJ), and incubated for 15-20 min in blocking solution (in a glass tray), with gentle agitation. The blocking solution comprises 325 ml 1-methyl-2-pyrrolidone (100 ml); 6 g succinic anhydride (1.8 g) and 15 ml sodium borate (pH 8.0) (4.6 ml). The slides are washed for 2 min. in H_2O (95°C); washed for 1 min. in 96% ethanol; and dried for 5 min. by centrifugation in a tabletop centrifuge at 1000 rpm.

The pellet, which must be clearly stained, is dissolved in in 18 µl H_2O for Klenow target or in 15 µl H_2O for Mirus target.

8. Hybridisation

The denaturation of labeled target is carried out by adding 1.5 µl denaturation buffer D1 (3M NaOH), after which the mixture is kept at room temperature for 5 min, and then place on iced, upon which 1.5 µl neutralisation buffer N1 (1M Tris pH 7.3, 3M HCl) is added.

Subsequently, 18 µl 2x hybridization buffer (preheated), comprising 4x SSC, 5x Denhardt, 0.5% SDS, is added at 60°C, after which the hybridization is started by, with a pipet, adding 30 µl target solution to the slide, next to the array.
A the cover glass (24x50 mm) is placed in position (without air bubbles), and the slides are incubated (o/n) at 45°C in a single incubation chamber (in which case 2 drops 10 µl 3x SSC are added) or in a large incubatotion tank containing water.

The hybridizations are washed by rinsing with 4x SSC 0.1% SDS (45°C); incubating for about 5 min in 2x SSC 0.1% SDS (45°C); incubating for 5 min 1x SSC 0.1% SDS (RT); incubating for 5 min in 0.5x SSC 0.1% SDS (RT); incubate for 2 min in 0.5x SSC (RT); followed by centrifugation for 10 min (500 rpm) in a tabletop centrifuge.

9. Scanning

The arrays are scanned using the Genetac 1000 scanner (Genomic solutions). The arrays are irradiated with a Xenon-lamp and the signals are detected using a CCD-camera. Filters for Cy-3 and Cy-5 are used. The scantime is about 180 sec.

EXAMPLE VI: Detection of AFLP markers on microarrays.

EXAMPLE VI-1: Detection of a mixture of 5 rice +2/+3 AFLP markers on an array containing 20 rice +2/+3 AFLP markers.

An array of 20 rice +2/+3 AFLP markers (probes) was prepared from cloned

The of sequences the +2/+3 AFLP primers used to generate these 20 AFLP markers are:

EcoRI	E11:	5'-GACTGCGTACCAATTCAA-3'	(SEQ ID no.12)
MseI	M47:	5'-GATGAGTCCTGAGTAACAA-3'	(SEQ ID no.13)
	M49:	5'-GATGAGTCCTGAGTAACAG-3'	(SEQ ID no.14)
	M50:	5'-GATGAGTCCTGAGTAACAT-3'	(SEQ ID no.15)

The AFLP reactions used to isolate the 20 AFLP +2/+3 makers were generated and resolved on sequence gels using the standard procedure (Vos et al., Nucleic Acids Research 23; 4407-4414, 1995 and EP 0 534 858). The AFLP markers were excised from a sequencing gel after transfer to Whatmann paper, followed by drying and exposure to X-ray film to visualize the fingerprint pattern and reamplified using primers :
5'-AGCGGATAACAATTTCACACAGGATAGACTGCGTACGAATTCA-3' (SEQ ID no.16) and
5'-CGCCAGGGTTTTCCCAGTCACGACGATGAGTCCTGATTAAC-3' (SEQ ID no.17) as described in the protocol.

After cutting with *Eco*RI and *Mse*I and purification using Qiagen PCR purification kits (Qiagen) the restricted AFLP marker fragments were cloned in plasmid vector digested with *Eco*RI and *Nde*I. After transformation to *E.coli*, recombinant clones were validated for the correct size insert by AFLP fingerprint analysis of pooled amplified clone inserts. The inserts of clones with validated inserts were sequenced using a standard dye terminator cycle sequencing kit (ABI) according to standard protocols supplied by the manufacturer.

Insert DNAs of individual validated clones were amplified from bacterial stocks by PCR using either unlabelled vector primers or Cy3-labelled vector primers as described (see protocol enclosed) and the PCR reactions were precipated using n-propanol and sodiumbicarbonate according to standard procedures. DNAs were resuspended in 50% DMSO to a final concentation of around 500 nanograms per microliter.

Microarrays were prepared by depositing in duplo around 250 picoliters (1

Following washing, images were taken at the Cy3 (Figure 2B) and Cy5 (Figure 2C) channels as described in Example VI-1 and the two images were superimposed electronically (Figure 2A). Figure 2 is annotated to facilitate interpretation of the spotting pattern and shows:

1) uniform deposition of all AFLP probes (red, green or yellow signal of all probes on false-color image after hybridization).
2) specific hybridization of the IR20 and 6383 AFLP markers 6, 8, 10, 14, 16, 18 and 20; Cy5, green signal on false-color image hybridized to unlabelled probes and green/yellow signal on false-color image hybridized to Cy3 labelled probes).
3) no hybridization of to the remaining AFLP markers 2, 4 and 12; Cy3, red signal on false-color image).

EXAMPLE VI-3. Detection of rice AFLP markers amplified in a +2/+2 AFLP reaction on an array containing 20 rice +2/+3 AFLP markers.

An array containing 20 rice AFLP markers (probes) was prepared exactly as described in Example VI-1. The array was processed as described and hybridized using a Cy5-labelled AFLP +2/+2 reaction (target) derived from parental line 6383, prepared with AFLP primers E11 and M15: 5'-GATGAGTCCTGAGTAACA-3' (SEQ ID no.18). This parental line is known to contain AFLP markers with names 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 as described in Example VI-1. The array was washed according to the conditions as described in Example VI-1 or protocols refered to in Example VI-1.

Following washing, images were taken at the Cy3 (Figure 3B) and Cy5 (Figure 3C) channels as described in EXAMPLE VI-1 and the two images were superimposed electronically (Figure 3A). The superimposed image is shown in Figure 3A with annotation to facilitate interpretation of the spotting pattern. Figure 3 shows:

1) uniform deposition of all AFLP probes (red, green or yellow on false-color image).
2) specific hybridization to the AFLP markers in positions 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 (Cy5; green signal on false-color image; in combination with Cy3

4) Hybridization of both IR20 and 6383 AFLP markers 2, 3, 12, 13 and 18; yellow on false-color image).

5) No hybridization to probes 5, 8, 16 and 19.

EXAMPLE VI-5. Detection of maize +2/+3 AFLP markers on an array containing 48 maize +3/+3 AFLP markers

An array of 48 maize +2/+3 AFLP markers (probes) was prepared from cloned AFLP markers generated using restriction enzymes *Eco*RI and *Mse*I and parental lines B73, Mo17, F2, Co255, DK105 and A7. The AFLP marker name, AFLP primer combination (PC) used, estimated mobility (size in basepairs) and the parental origin of these 48 AFLP markers are:

AFLP Marker Name	PC	Size (bp)	Parent Line
A1	E33/M50	596	Mo17
A2	E33/M50	588	Mo17,Co255
A3	E33/M50	580	B73,A7
A4	E33/M50	566	B73,Mo17,Co255,A7
A5	E33/M50	526	B73,Mo17,Co255
A6	E33/M50	503	F2,DK105
A7	E33/M50	459	DK105
A8	E33/M50	453	B73,DK105
A9	E33/M50	447	B73,Mo17,Co255,DK105,A7
A10	E33/M50	434	Mo17,F2,Co255,
A11	E33/M50	424	F2,Co255,A7
A12	E33/M50	416	B73,Mo17,F2,Co255,DK105,A7
C1	E33/M50	308	Mo17,F2,Co255
C2	E33/M50	304	B73,Mo17,F2,Co255,DK105,A7
C3	E33/M50	292	B73,Mo17,F2,DK105,A7
C4	E33/M50	290	B73,Mo17,F2,Co255,DK105,A7

G11	E33/M50	106	B73,Mo17,F2,DK105,A7
G12	E33/M50	103	B73,F2,Co255,DK105,A7

The sequences the +2/+3 AFLP primers used to generate these 48 AFLP markers are:

E33: *EcoRI*: 5' GACTGCGTACCAATTCAAG-3' (SEQ ID no.19)

M50: *MseI*: 5'GATGAGTCCTGAGTAACAT-3' (SEQ ID no.20)

The AFLP reactions used to isolate the 48 AFLP +3/+3 makers were generated, excised, reamplified, purified, cloned and validated as described in the protocol of Example VI-1. The inserts of clones with validated inserts were sequenced using a standard dye terminator cycle sequencing kit (ABI) according to standard protocols supplied by the manufacturer.

Insert DNAs of individual validated clones were amplified from bacterial stocks by PCR using either unlabelled vector primers or Cy3-labelled vector primers as described in Example VI-1.

PCR reactions were precipitated and dissolved as described in the protocol of Example VI-1. Microarrays were prepared by depositing in duplo around 250 picoliters (1 time spotting) or 1250 picoliters (5 times spotting) of either the unlabelled or the Cy3-labelled DNA solutions, processed and hybridized according to the protocol of Example VI-1. The target was a mixture of complete +2/+3 E33/M50 AFLP reactions of the parental lines B73 and F2, after labeling the B73 DNA with Cy5 dye (Amersham Pharmacia Biotec), and the F2 DNA with Cy3, by Klenow enzyme according to standard procedures (see protocol enclosed). After washing according to the protocol, the slide was scanned at the Cy5 and Cy3 channels for 180 seconds each using a Genetac1000 microarray slide scanner (Genomic Solutions, Ann Arbor, MI).

The superimposed image of both channels is shown in Figure 5 with annotation to facilitate interpretation of the spotting pattern. Figure 5 shows:

1) uniform deposition of all AFLP probes (red, green or yellow signal of all probes on false-color image after hybridization).

9. C7	E11/M62	171	Columbia.
10. C9	E11/M62	163	Columbia.
11. C11	E11/M62	153	Columbia.

The sequences of the +2/+3 AFLP primers used to generate these 11 AFLP markers are:

E11: *EcoRI*: 5'-GACTGCGTACCAATTCAA-3' (SEQ ID no.21)

M62: *MseI*: 5'-GATGAGTCCTGAGTAACTT-3' (SEQ ID no.22)

The method used to generate the 11 AFLP +2/+3 markers and the preparation and processing of the arrays containing these 11 Arabidopsis AFLP markers is as described in Example VI-1 or protocols refered to in Example VI-1.

The arrays were hybridized with targets consisting of a Cy5-labelled AFLP +2/+3 reaction derived Colombia or Landsberg erecta, which were prepared as described in Example I. The AFLP used to generated the labelled target were E11: 5'-GACTGCGTACCAATTCAA-3' (SEQ ID no.23)
and M62: 5'-GATGAGTCCTGAGTAACTT-3' (SEQ ID no.24). With this primer combination, the parental line Columbia is known to contain the AFLP markers A3, A5, C1, C3, C7, C9 and C11 and parental line *Landsberg erecta* is known to contain the AFLP markers A7, A9, A11 and C5.

The array was washed according to the conditions discribed in Example VI-1 or protocols refered to in Example VI-1. Folowing washing, images of the array were taken at the Cy3 and Cy5 channels with a 180 second exposure time for both channels and the images were superimposed, as described in Example VI-1 (Figure 6). Figure 6A shows:

1) Specific hybridization of the Columbia AFLP markers A3, A5, C1, C3, C7, C9 and C11 (green signals on false-color image).
2) An anonymous Cy3-labeled AFLP fragment at position A1 which marks the start position of the array (red signal on false-color image)

Figure 6B shows:

1) Specific hybridization of the *Landsberg erecta* AFLP markers A7, A9, A11 and C5, (green signal on false-color image).

19.	H5	E01/M17	131	52201
20.	H9	E01/M17	114	52201
21.	H11	E01/M17	103	52201

The sequences of the +1/+2 AFLP primers used to generate these cDNA-AFLP fragments are:

*Eco*RI: E01: 5'-GACTGCGTACCAATTCA-3' (SEQ ID no. 25)

*Mse*I: M16: 5'-GATGAGTCCTGAGTAACC-3' (SEQ ID no.26)

*Mse*I: M17: 5'-GATGAGTCCTGAGTAACG-3' (SEQ ID no.27)

The cDNA-AFLP reactions used to isolate the 21 +1/+2 fragments were generated and resolved on sequence gels using the standard procedure. Arrays were prepared according to the procedures described in EXAMPLE VI-1. cDNA-AFLP fragments were spotted in duplo as described in Example VI-1.

The slides were processed according to standard procedures (see protocol enclosed) and hybridised overnight at 45°C temperature with Cy3-labeled +2/+3 AFLP reactions (targets) of the following six tomato lines:

1. *Lycopersicon Esculentum* (L. esc.) accession Moneyberg
2. *L. Peruvianum* accession LA1708
3. *L. Hirsutum* G1209
4. *L. Chmielevski* LA1848
5. *L. Pimpinellifolium* LA722
6. *L .Pennelli* LA716

The sequences of the AFLP primers involved are E12 (5'-GACTGCGTACCAATTCAC-3', SEQ ID no. 28) and M16 (sequence given above, SEQ ID no.26). Labelling with Cy3 dye (Amersham Pharmacia Biotec) by Klenow enzyme was carried out according to standard procedures as described in Example VI-1. After washing according to the protocol, the slides were scanned at the Cy3 channel in the automatic exposure mode using a Genetac1000 microarray slide scanner. The images of all six hybridisations are shown in Figure 7 (A-F) with

99f69	5'-GGCAATGCAAGTAGATACTTC-3'	(SEQ ID no.35)
99f70	5'-GAAGTATCTACTTGCATTGCC-3'	(SEQ ID no.36)
99f19	5'-CAGTGTGCTAGTTGATTCCAG-3'	(SEQ ID no.37)
99f20	5'-CTGGAATCAACTAGCACACTG-3'	(SEQ ID no.38)

The array was processed as described and hybridized with a target, consisting of a mixture of equal volumes of Cy5-labelled AFLP +2/+3 reactions (target) derived from the parental lines IR20 and 6383, prepared with AFLP primers E11 and M49. Thus in the labeled target one of the strands of the AFLP +2/+3 reaction fragments is labeled with Cy5. The mixture of parental lines IR20 and 6383 is known to contain AFLP markers 6, 8 and 10 as described in Example VI-1. The array was washed according to the conditions as described in Example VI-1 or protocols referred to in Example VI-1.

Following washing, images were taken at the Cy3 (Figure 8B) and Cy5 (Figure 8C) channels both at 180 seconds exposure time and the two images were superimposed electronically (Figure 8A). The superimposed image of both channels is shown in Figure 8A with annotation to facilitate interpretation of the spotting pattern. Figure 8 shows:

1) uniform deposition of all AFLP probes (red, green or yellow on false-color image after hybridization).
2) specific hybridization to AFLP markers 6, 8 and 10; Cy5, green on false-color image).
3) specific hybridization to reverse sequence oligo's corresponding to the unlabeled strand of the AFLP markers 6, 8 and 10; Cy5, green on false-color image).
4) No hybridisation to the forward sequence oligo's corresponding to the labeled strand of the AFLP markers 6, 8 and 10.
5) No hybridization to AFLP markers or oligo's corresponding to AFLP markers 2 and 4.

6. Array according to any of the preceding claims, in which the restriction fragments have been derived from genomic DNA, and in which at least 50%, preferably at least 70%, more preferably at least 90%, of the nucleic acid sequences bound to the carrier comprise the sequence of a restriction fragment that corresponds to an AFLP-marker.

7. Array according to claim 6, comprising a plurality of AFLP-markers taken from a single individual or from a group of related individuals.

8. Array according to claim 6 or 7, comprising several sets of AFLP-markers, in which each set comprises one or more markers taken from a single individual, wherein said sets of one or more markers have been taken from individuals belonging to a group of related individuals.

9. Array according to any of claims 6-8, in which the AFLP-markers have been taken from individuals belonging to the same species of plant, animal or micro-organism.

10. Array according to any of claims 7-9, in which the AFLP-markers have been taken from, or are representative for, different subspecies, varieties, cultivars or races of the same species.

11. Array according to claim 9 or 10, in which the AFLP-markers have been taken from plants including but not limited to wheat, barley, maize, tomato, pepper, lettuce or rice.

12. Array according to any of claims 6-9, in which the AFLP-markers have been taken from the human genome.

13. Array according to any of claims 1-12, and in particular according to claim

18. Method according to claim 17, comprising the steps of:

a) analysing at least one cDNA using AFLP-methodology to provide a cDNA-AFLP fingerprint, said fingerprint comprising at least one, and usually a plurality, of bands;

b) isolating from at least one of said bands at least one nucleic acid sequence;

c) optionally further amplifying, purifying and/or modifying the nucleic acid sequence;

d) attaching the nucleic acid sequence to the carrier; and

e) repeating steps a) to d) for different bands and/or for different cDNAs to build up an array.

19. Array, obtainable by the method of any of claims 15-18.

20. Method for analysing a nucleic acid sequence or a mixture of nucleic acids sequences, comprising contacting said nucleic acid or mixture under hybridizing conditions with an array according to any of claims 1-14 or 19 or obtained according to the method of any of claims 15-18.

21. Method according to claim 20, in which the nucleic acid sequence or mixture is suspected to comprise at least one sequence that corresponds to a restriction fragment sequence present in the nucleic acid sequences present in the array, more specifically to an AFLP-marker present in the array.

22. Method according to claim 20 or 21, in which the nucleic acid sequence or mixture comprises DNA, in particular genomic DNA or a mixture of restriction fragments derived from genomic DNA.

23. Method according to any of claims 20-22, in which the nucleic acid sequence or mixture comprises a mixture of restriction fragments obtained by restricting a genomic DNA with the same frequent cutter restriction enzyme and rare cutter

30. Non-processed or processed results or data obtainable by analysing a nucleic acid or mixture of nucleic acids with an array according to any of claims 1-14 or 19, or by a method according to any of claims 20-28, for instance in the form of an image, of a score, of digital or analog data, optionally stored on a suitable data carrier, including paper, photographic film, computer disc or files, or as a database.

Fig 2a

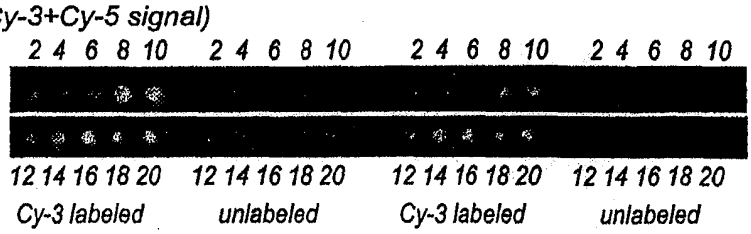


Fig 2b

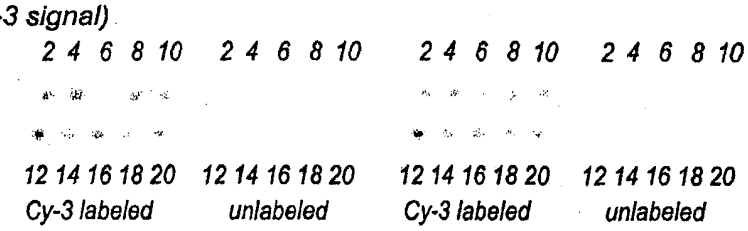


Fig 2c

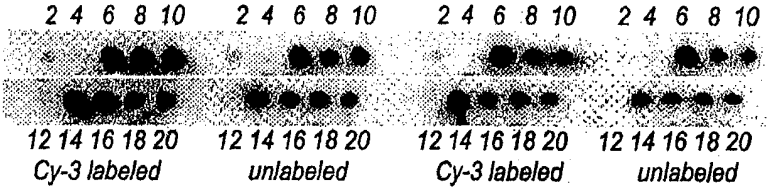


Fig 5

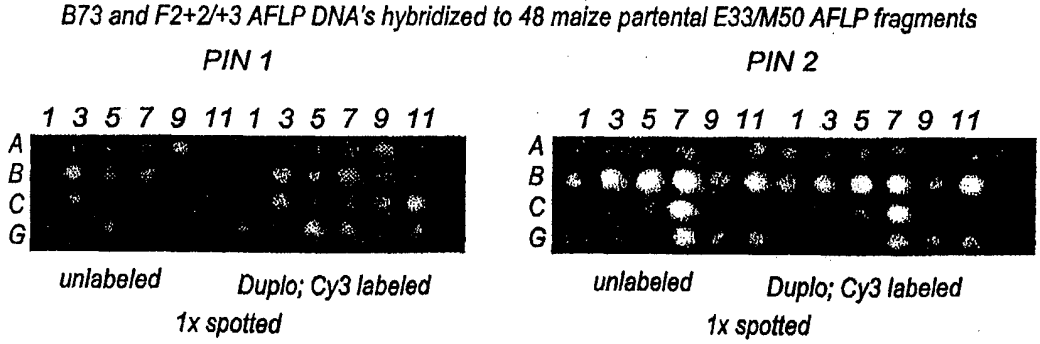


Fig 6a

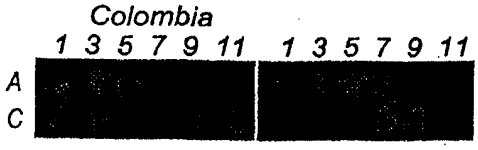


Fig 6b

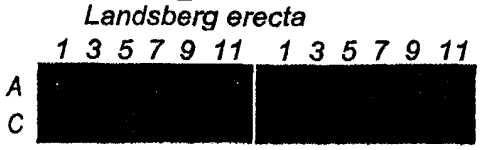


Fig 4a (Cy-3+Cy-5 signal)

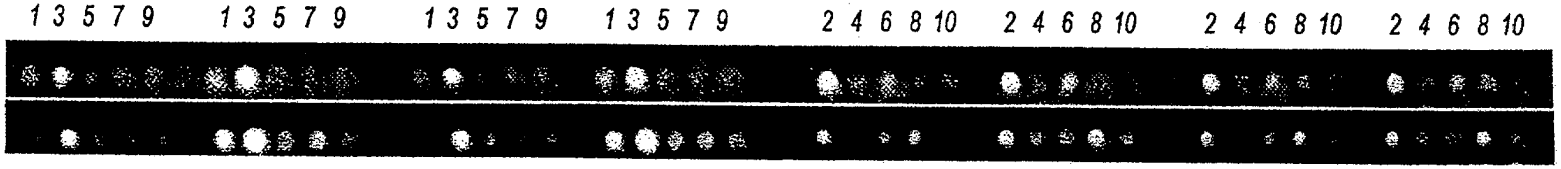


Fig 4b (Cy-3 signal)

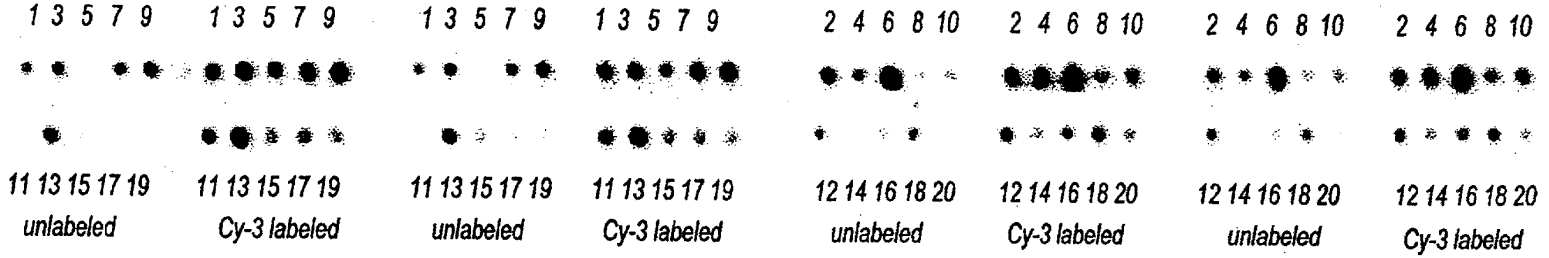


Fig 4c (Cy-5 signal)

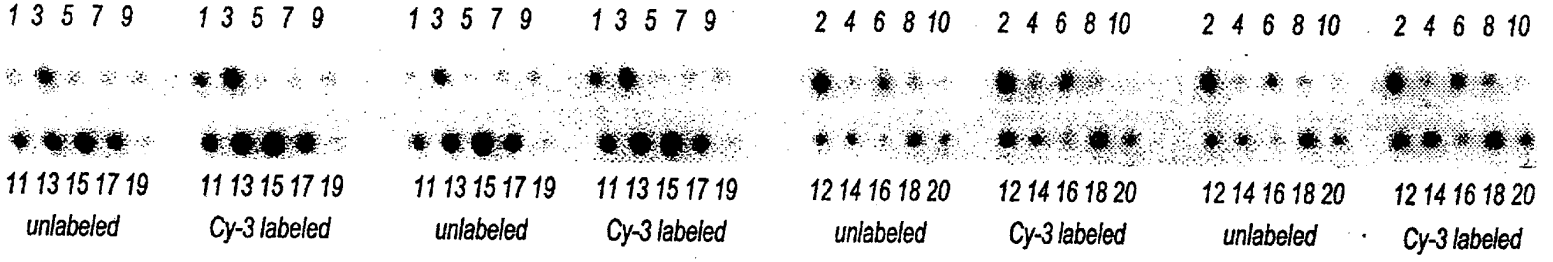


Fig 8a

(Cy-3+Cy-5 signal)

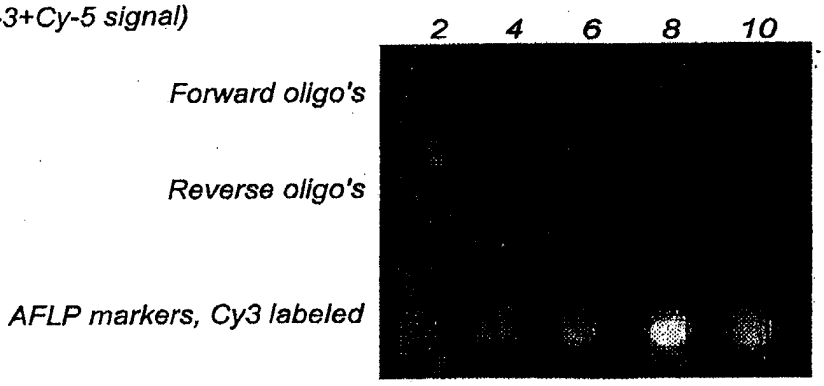


Fig 8b

(Cy-3 signal)

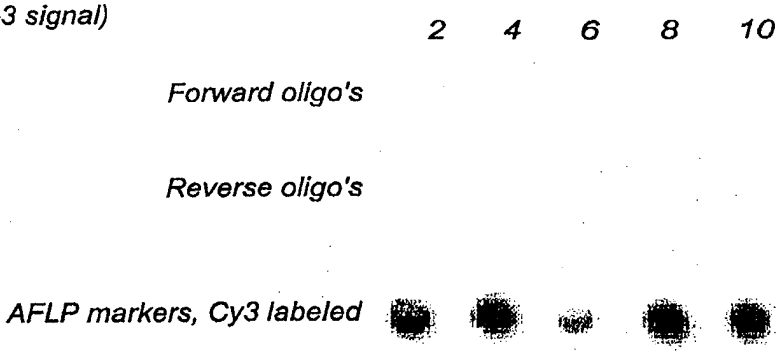


Fig 8c

(Cy-5 signal)

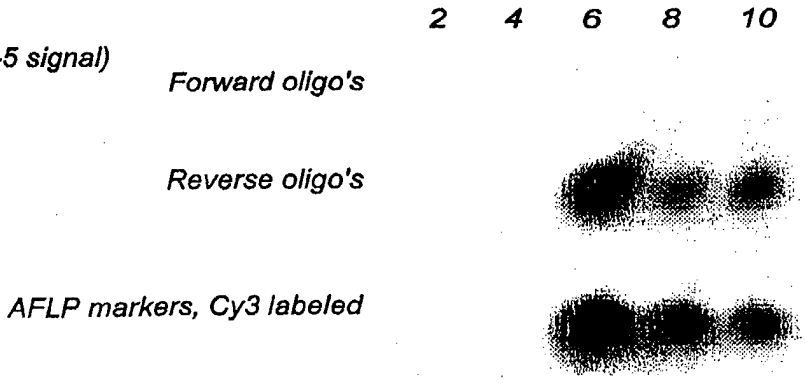


Fig 10

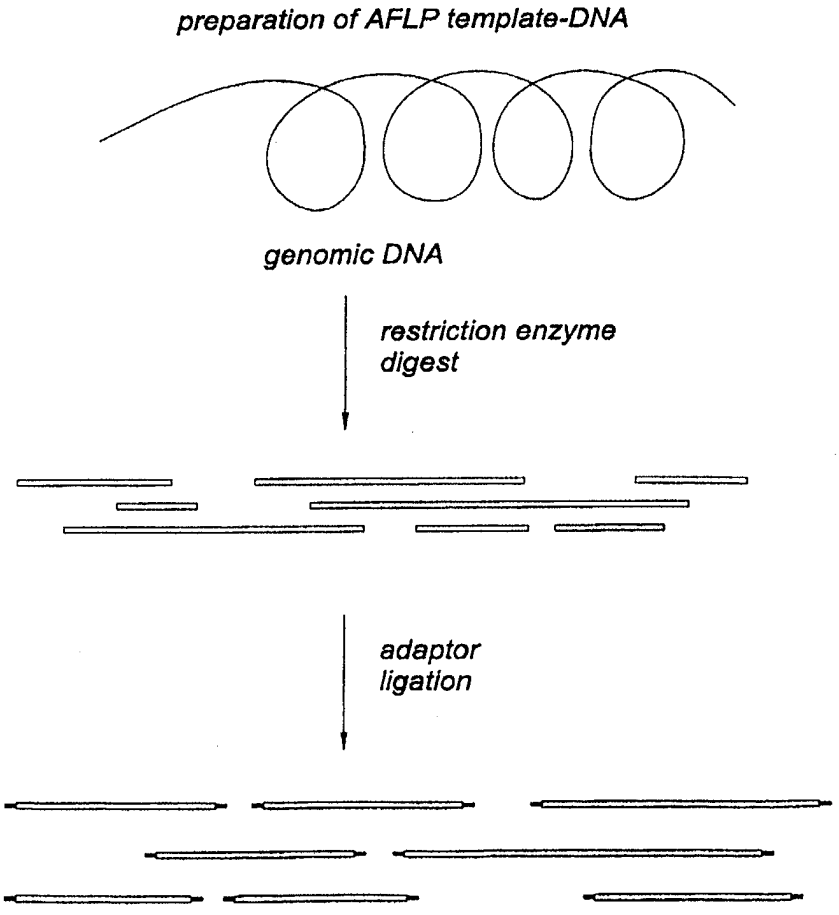
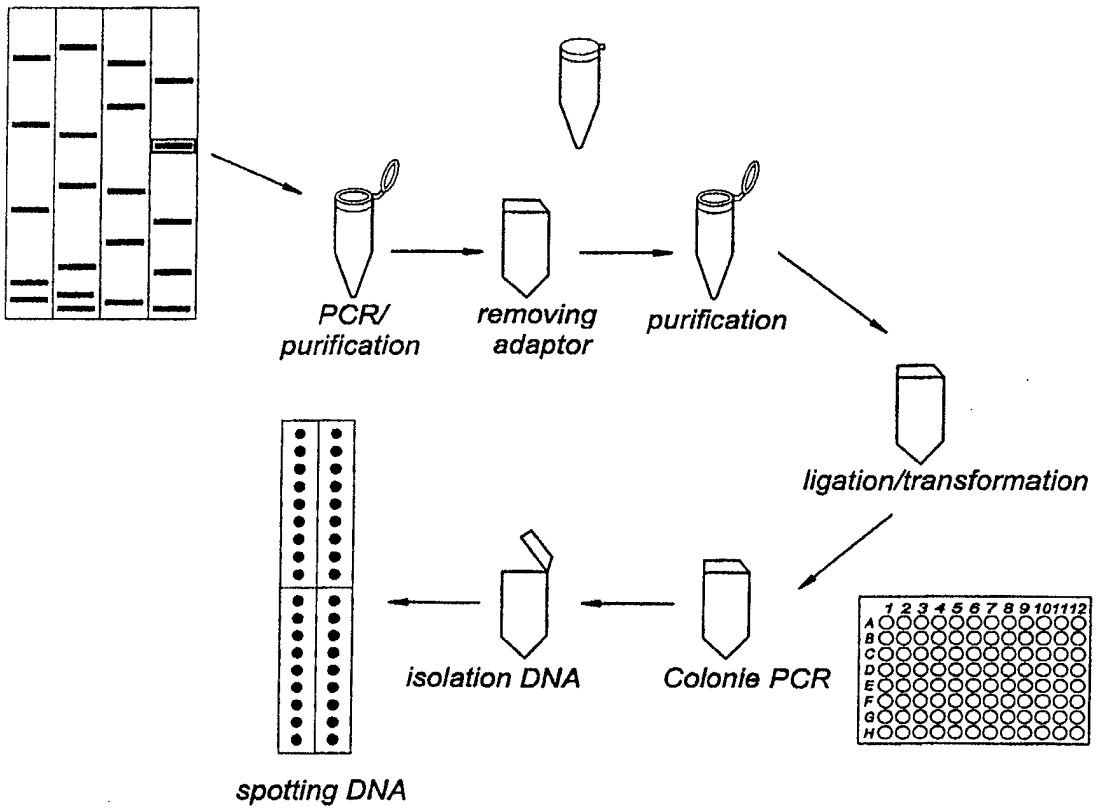

preparation of AFLP template-DNA
genomic DNA
restriction enzyme digest
adaptor ligation

Fig 12


PCR/
purification
removing
adaptor
purification
ligation/transformation
Colonie PCR
isolation DNA
spotting DNA
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H

INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/NL 99/00743

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	VOS P ET AL: "AFLP: a new technique for DNA fingerprinting" NUCLEIC ACIDS RESEARCH, 11 November 1995 (1995-11-11), XP002086756 cited in the application the whole document	
P,X	WO 99 23256 A (LUCITO ROBERT ;WIGLER MICHAEL (US); COLD SPRING HARBOR LAB (US)) 14 May 1999 (1999-05-14) see whole doc. esp. claims	1-30

Form PCT/ISA/210 (continuation of second sheet) (July 1992)